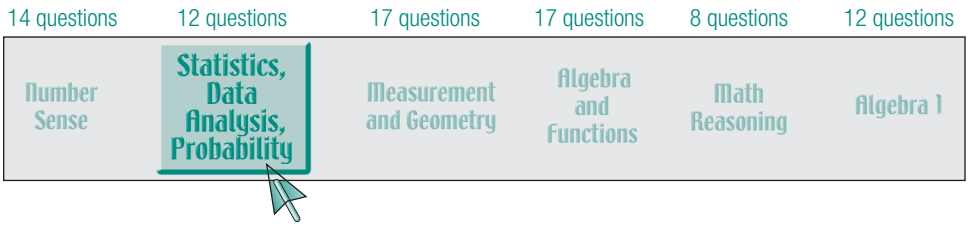
2

Statistics, Data Analysis, and Probability Strand



Twelve of the 80 CAHSEE multiple-choice questions are based on seven selected standards from the Statistics, Data Analysis, and Probability strand for grades six and seven.

WHAT DO THE STATISTICS, DATA ANALYSIS, AND PROBABILITY STANDARDS ASK ME TO DO?

The CAHSEE questions from Statistics, Data Analysis, and Probability strand will ask you to:

- understand data displays including bar graphs and scatterplots
- find the mean, median, and mode of a data set
- express the probability of an event as a ratio, a decimal, or a percent
- know whether events are independent or dependent.

Vocabulary

The words below have appeared previously on the CAHSEE. If any of these words are unfamiliar to you, look them up in the CAHSEE Math Vocabulary list in the appendix at the back of this study guide, or check with your math teacher.

bar graph	scatterplot	probability
pie chart	correlation	random
	mean	independent events
	median	dependent events

WHY ARE STATISTICS, DATA ANALYSIS, AND PROBABILITY IMPORTANT?

Many occupations require a basic working knowledge of statistics and probability because vast amounts of data can now be gathered electronically and analyzed by computers. Even though you may not have to display the data, design the poll, or calculate the probabilities yourself, you will need to understand statistical information to make intelligent business decisions and healthy lifestyle choices. The anchor problem *Shipping Motorcycle Parts* illustrates how several of the statistics and probability standards are useful in a business situation.

But before we try *Shipping Motorcycle Parts,* let's first look at some released CAHSEE questions, with answers, for this strand.

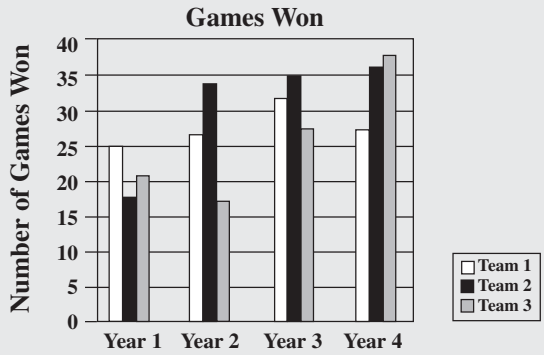
HOW WILL THE CAHSEE TEST MY KNOWLEDGE OF STATISTICS, DATA ANALYSIS, AND PROBABILITY?

The CAHSEE tests five of the 14 grade six standards and all three of the grade seven standards from the Statistics, Data Analysis, and Probability strand. Let's start by looking at four of these standards and some actual CAHSEE questions based on them. Each box that follows contains one of the standards, a released question based on that standard, and a solution with explanation.

P6 2.5 Identify claims based on statistical data and, in simple cases, evaluate the validity of the claims. [1 question]

Released CAHSEE Question

The number of games won over four years for three teams is shown on the graph below.



Which statement is true based on this information?

A Team 3 always came in second.
B Team 1 had the best average overall.
C Team 1 always won more games than Team 3.
D Team 2 won more games each year than in the previous year.

M10300

Solution

To find the number of games a team won, find the top of the team's bar, then see which number the top of the bar is aligned with on the vertical line, or *y*-axis, labeled "Number of Games Won." For example, in Year 1, you can see that Team 1 won 25 games, because the top of the bar is aligned with the number 25 along the *y*-axis.

By reading the bar graph, you can see that Option A is not true because Team 3 came in second only in Year 1. To determine whether Option B is true, you must calculate the overall average, or mean, for each team. To find the mean for Team 1, you add the number of games won in each of the four years ($25 + 27 + 32 + 28 = 112$); then divide by the number of years ($112 \div 4 = 28$). Using this same method, you find that the overall average for Team 2 is 30.5 ($122 \div 4$) and the overall average for Team 3 is 26 ($104 \div 4$). Team 1 did not have the best overall average, so Option B is not true. Option C is not true because Team 3 won more games than Team 1 in Year 4. Option D is a true statement because Team 2 won more games each year than in the previous year. So the correct answer is **D**.

P6 3.1 Represent all possible outcomes for compound events in an organized way (e.g., tables, grids, tree diagrams) and express the theoretical probability of each outcome. [1 question]

Released CAHSEE Question

To get home from work, Curtis must get on one of the three highways that leave the city. He then has a choice of four different roads that lead to his house. In the diagram below, each letter represents a highway, and each number represents a road.

Road	Highway A	Highway B	Highway C
1	A1	B1	C1
2	A2	B2	C2
3	A3	B3	C3
4	A4	B4	C4

If Curtis randomly chooses a route to travel home, what is the probability that he will travel Highway B and Road 4?

A $\frac{1}{16}$

B $\frac{1}{12}$

C $\frac{1}{4}$

D $\frac{1}{3}$

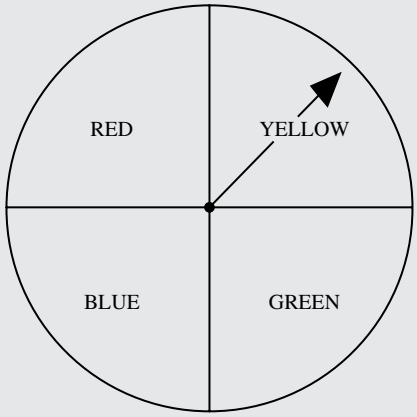
M02512

Solution

The chart gives you an organized presentation of the 12 possible routes that Curtis can follow. Because Curtis chooses his route randomly, each of the twelve possible routes shown in the chart is equally likely. Traveling Highway B and Road 4 is only one of these 12 equally likely possibilities. Therefore, the correct answer is **B**, $\frac{1}{12}$.

P6 3.3 Represent probabilities as ratios, proportions, decimals between 0 and 1, and percentages between 0 and 100 and verify that the probabilities computed are reasonable; know that if *P* is the probability of an event, 1-*P* is the probability of an event not occurring. [2 questions]

Released CAHSEE Question



The spinner shown above is fair. What is the probability that the spinner will <u>not</u> stop on red if you spin it one time?

A $\frac{1}{4}$

B $\frac{1}{3}$

C $\frac{3}{4}$

D $\frac{4}{3}$

M00094

Solution

Because the spinner is "fair," this means that the four outcomes—"red," "yellow," "blue," and "green"—are each equally likely to be the result of a spin. Because three of the four possibilities are not "red," the probability of not spinning "red" is answer **C**, $\frac{3}{4}$.

P6 3.5 Understand the difference between independent and dependent events. [1 question]

Released CAHSEE Question

Heather flipped a coin five times, and each time it came up heads. If Heather flips the coin one more time, what is the theoretical probability that it will come up tails?

A $\frac{1}{6}$

B $\frac{1}{2}$

C $\frac{3}{5}$

D $\frac{5}{6}$

M02171

Solution

We assume this coin is fair, which means that on any particular coin flip, the chance of getting a head or getting a tail are equally likely. For a truly fair coin, the flips that came before have no effect on the outcome of the next flip; each flip of the coin is *independent* of all the flips that came before. So, the probability of getting tails on the sixth flip is still $\frac{1}{2}$, choice **B**.

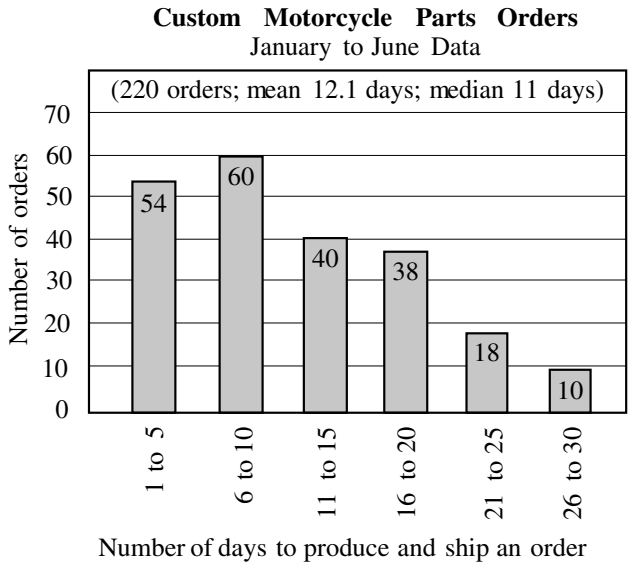
USING STATISTICS, DATA ANALYSIS, AND PROBABILITY STANDARDS IN A REAL-LIFE SITUATION

You might again be asking yourself "When will I ever need to use this stuff?" To help you get the "big picture," the remaining three Statistics, Data Analysis, and Probability standards will be illustrated by a real-life problem that a person might encounter after high school: *Shipping Motorcycle Parts.* Even though the CAHSEE doesn't include problems like this one, you might find it easier to remember one large problem (an "anchor problem") where many of the skills are combined, rather than trying to recall how to do each of the standards individually.

Shipping Motorcycle Parts, Part I

You work as a sales agent for Custom Motorcycle Parts Unlimited. When you make a sale, you tell your customers that their orders will be shipped "in about 10 working days." But lately you are getting a lot of calls back from your customers who haven't received their orders on time.

You call the production manager and ask how long it takes her employees to ship a typical order after they receive it. Because their computerized system collects data on how long it takes to make and send out each order, the production manager is able to supply you with a bar graph showing how many days it took to get orders out during the past six months:



The production manager also gives you the data for the ten orders shipped last week, the first week of July. The number of days to produce and ship each of the ten orders sent out last week were:

6, 15, 4, 19, 10, 21, 4, 17, 24, 20

Based on the data you have, what would you tell your customers about how many days it will take to produce and ship out their orders?

Before turning the page, find the mean and median of last week's data.

Compare your answers with the six-month data shown in the bar graph.

Shipping Motorcycle Parts, Part 1 Solution

The bar graph and last week's data give you the information you'll need so you can be more helpful to your customers.

First let's look at the data for last week:

6, 15, 4, 19, 10, 21, 4, 17, 24, 20

P6 1.1 Compute the ~~range~~, mean, and median and mode of data sets. [3 questions]. (Note: The crossed out portion will not be tested on the CAHSEE.)

What was the average number of days for shipping an order last week? For this data, the average could be either the *mean* or the *median.*

To find the *mean* we need to find the sum of the data and then divide the sum by the number of data items.

$$6 + 15 + 4 + 19 + 10 + 21 + 4 + 17 + 24 + 20 = 140$$

Since there are ten data items we divide the sum by 10: $140/10 = 14$ days. So the mean number of days to produce an order for those shipped last week was 14 days.

To find the *median* of a data set, we start by putting all numbers in order. Last week's data, written in order from least to greatest is

4, 4, 6, 10, 15, 17, 19, 20, 21, 24

For an odd number of data, the median is the middle number; for an even number of data, it's the average of the middle two numbers. In this case we have ten data items, an even number. The middle two numbers in the set are 15 and 17. So the median is the average of 15 and 17, which is 16 days.

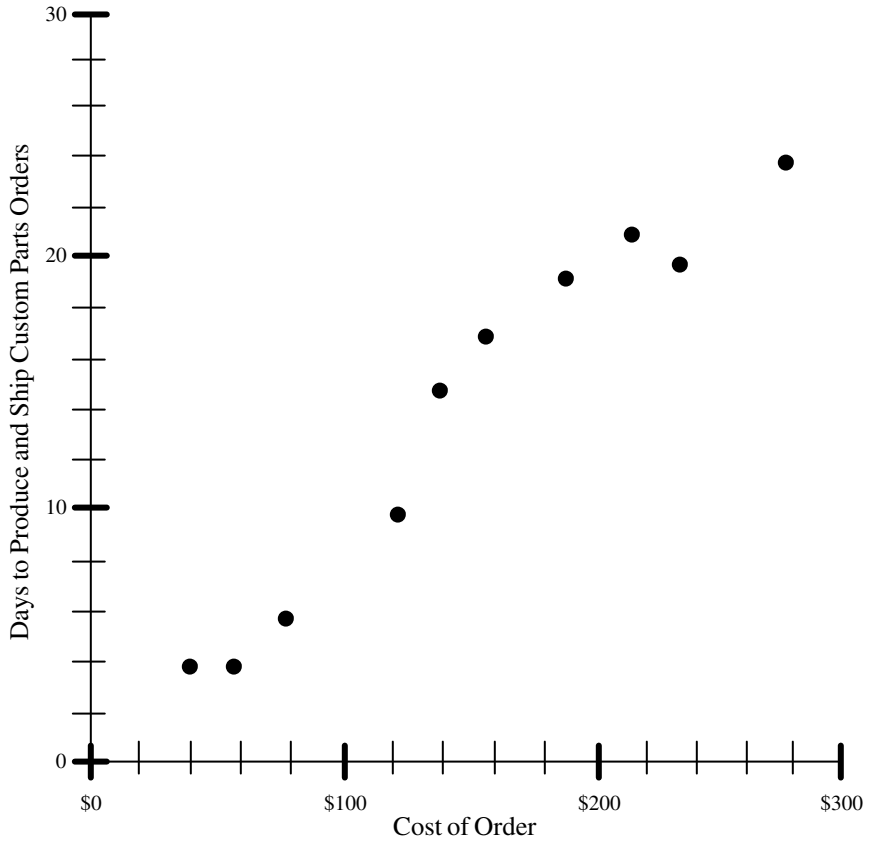
P7 1.1 Know various forms of display for data sets, ~~including a stem-and-leaf plot or box-and-whisker plot~~; use the forms to display a single set of data or to compare two sets of data. [2 questions] (Note: The crossed out portion will not be tested on the CAHSEE.)

Now look back at the bar graph that shows the data for the past six months. How does last week's data compare with the data in the bar graph? The mean of last week's data was 14 days, which is greater than the 12.1 day mean for the past six months. Last week's median of 16 days was also greater than the six month median of 11 days. The bar graph displays the distribution of the data. The bars on the graph show that almost half the orders are shipped out within 10 days, but 28 of the orders took 21 days or more.

So what should you tell your customers about shipping times? Based on this data, you might want to tell your customers to expect their order to be shipped in about 12 to 14 days, but some orders may take up to 30 days before shipping.

Shipping Motorcycle Parts, Part 2

You would like to give your customers more precise information about shipping times. From your experience you know that expensive orders take longer to produce and ship. So you get additional data about the price for each order shipped last week and make a scatterplot of the paired data.



Does the scatterplot show a pattern? What does it tell you about the relationship between the cost of an order and the days it takes to send it out?

Based on the trend indicated by the scatterplot, about how many days will it take to produce and ship a $100 order? About how long for a $200 order?

P7 1.2 Represent two numerical variables on a scatterplot and informally describe how the data points are distributed and any apparent relationship that exists between the two variables (e.g., between time spent on homework and grade level). [2 questions]

Shipping Motorcycle Parts, Part 2 Solution

If you simply "eyeball" the trend in the scatterplot, it looks like a $100 order takes around seven or eight days to produce and ship; a $200 order takes about 18 to 20 days.

Now you are ready to test what you've learned by trying a few CAHSEE Released questions from the Statistics, Data Analysis, and Probability strand. Answer the questions in the next section—the Practice Test—and then check your answers using the answer key provided at the end.

(Note: The CAHSEE questions used as examples throughout this Study Guide are questions that were used on prior CAHSEEs. These items will not be used in future CAHSEEs.)

STATISTICS, DATA ANALYSIS, AND PROBABILITY PRACTICE TEST

1. Three-fourths of the 36 members of a club attended a meeting. Ten of those attending the meeting were female. Which one of the following questions can be answered with the information given?

A How many males are in the club?

B How many females are in the club?

C How many male members of the club attended the meeting?

D How many female members of the club did not attend the meeting?

M00261

2. Mr. Gulati is holding five cards numbered 1 through 5. He has asked five students to each randomly pick a card to see who goes first in a game. Whoever picks the card numbered 5 goes first. Juanita picks first, gets the card numbered 4, and keeps the card. What is the probability that Yoko will get the card numbered 5 if she picks second?

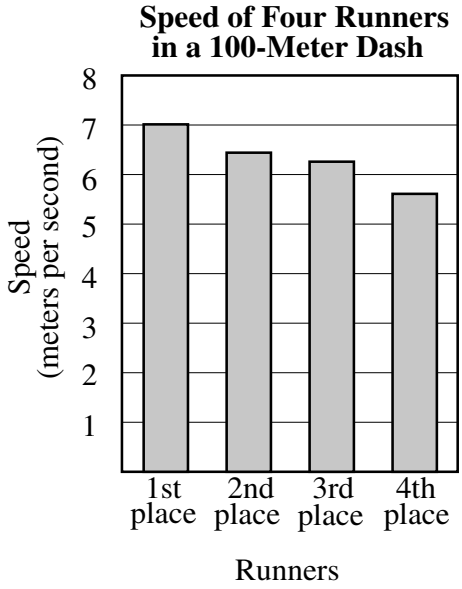
A $\frac{1}{2}$

B $\frac{1}{3}$

C $\frac{1}{4}$

D $\frac{1}{5}$

M02145



3. Based on the bar graph shown above, which of the following conclusions is true?

A Everyone ran faster than 6 meters per second.

B The best possible rate for the 100-meter dash is 5 meters per second.

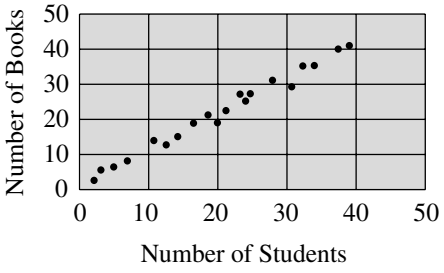
C The first-place runner was four times as fast as the fourth-place runner.

D The second-place and third-place runners were closest in time to one another.

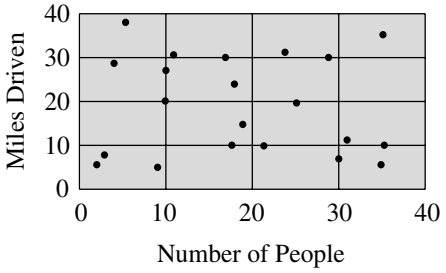
M00279

4. Which scatter plot shows a negative correlation?

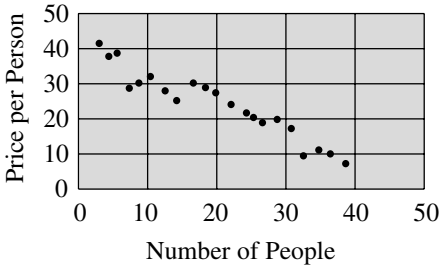
A



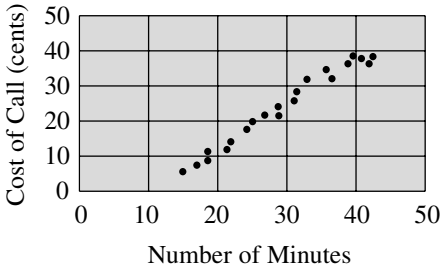
C



B



D



M02546

STATISTICS, DATA ANALYSIS, AND PROBABILITY ANSWER KEY

Question Number	Standard	Correct Answer
1	P6 2.5	C
2	P6 3.5	C
3	P7 1.1	D
4	P7 1.2	B